

Exploring what works in professional development: An assessment of a prototype intervention and its accompanying design principles.

Albert Zeggelaar^{a*}, Marjan Vermeulen^b, Wim Jochems^c

^aROC van Amsterdam, Amsterdam, The Netherlands; ^bWelten Institute, Open University, Heerlen, The Netherlands; ^cOpen University, Heerlen, The Netherlands

* corresponding author:

Albert Zeggelaar MSc
ROC Nijmegen, Campusbaan 6
6512 BT Nijmegen, The Netherlands
e-mail: a.zeggelaar@roc-nijmegen.nl
0031650243135

Prof. Dr. Marjan Vermeulen
Open University, Welten Institute
Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands
e-mail: Marjan.Vermeulen@ou.nl
003145 - 576 22 22

Prof. Dr. Wim Jochems
Open University, Welten Institute
Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands
e-mail: w.m.g.jochems@tue.nl
003145 - 576 22 22

Exploring what works in professional development: An assessment of a prototype intervention and its accompanying design principles.

The growing complexity and numerous changes in Dutch vocational education and training (VET) pressure teachers to continually develop their teaching profession. Therefore professional development (PD) is needed, but policy-initiated and imposed PD might be ineffective. Thus the purpose of this research is to evaluate which design requirements (DR) in a mandatory PD programme delivered in a VET institute appear to show the most effectiveness. Based on a review study aimed at detecting effectiveness enhancing DR an assessor PD programme was redesigned and implemented. This programme was evaluated with respect to effectiveness on five different levels of impact considering participants' attitudes, their reactions, what they had learned or what behaviour they developed, and organizational support. Each level of impact was evaluated using existing and self-constructed instruments including questionnaires, surveys, interviews and observations. Thirty-nine teachers took part in this redesigned PD programme. Several DR appeared to show measurable levels of effectiveness: tolerance for complex contexts, contribution to a professional identity and team based and collective participation.

Keywords: professional development; evaluation; design requirements; levels of impact; vocational education and training

Introduction

Like other educational sectors, vocational education and training (VET) can be characterized as complex (Volmari *et al.* 2009). Next to the substantial changes in the VET teacher profession, the professions the sector provides for are subjected to changes as a result of demographic developments, need for new and high skills in work, organization and labour market developments (Béduwé *et al.* 2009). This results in challenges keeping VET teachers ‘familiar with the fast-changing requirements of modern workplaces’ (OECD 2009, p.49). Furthermore because VET is seen as a large supplier of workforce, increased attention and importance of VET is placed in international policy agendas (Béduwé *et al.* 2009) resulting in output and performance management and accompanying cultures characterizing educational systems and sectors in most parts of the developed world (Sachs and Mockler 2012). Policy consequences to enhance the quality of VET is often based on the quality and professional development (PD) of the teacher being the ‘key element in overall quality control in VET institutions’ (OECD 2009, p. 48). Despite this general acceptance of PD being essential for the quality of education, incoherence between the goals of education, teachers’ PD, and the ‘politics of accountability’ (Ben-Peretz 2012, p.57) can prevent PD interventions, at the teaching professional level, from being implemented effectively. Therefore the central issue in this article is the enhancement of the quality of PD.

Theoretical background

In spite of an apparent consensus on the core features of effective professional development (Desimone 2011, Wayne *et al.* 2008), research frequently points out the ineffectiveness of various programmes (Guskey 2002, Borko 2004, Steiner 2004, Desimone 2009, Opfer and Pedder 2011). PD is often designed and offered to ensure that teachers are more compliant in delivering government initiated-policies more effectively and efficiently (Day 2012). This creates dilemmas for those teachers who professionalize for the benefit of academic, social and student development. For them, these dilemmas can cause inconsistency within the chain of centrally initiated policy, the offer in PD programmes and teachers’ conventions and beliefs. This constrains an effective implementation of PD in practice. Therefore, in order to evaluate the effectiveness of PD programmes, it is necessary to establish the purpose of PD (programmes) and understand the conditions and contexts in which PD is offered (Guskey 2000, Grollmann 2008).

Our research into the purpose of PD explains this inconsistency as tension caused by inconsistent expectations. For example, teacher quality is enhanced because teachers themselves actively engage in PD for the benefit of student achievement. However they fail to engage in PD to fulfil demanding policy agendas although both develop their professionalism leading to improved practice. In order to decrease this inconsistency conceptually Evans’ (2008, 2011) framework proved helpful as it harbours both perspectives. She explains the aims of PD as a concept built of three elements. Firstly, PD should focus on teachers’ ‘attitudinal development’ (Evans 2008, p. 16) which improves knowledge, and teachers’ behaviour focused on the improvement of student learning (see also: Guskey 2002a, Desimone *et al.* 2002, Villegas-Reimers 2003, Borko 2004, OECD 2005, Veen *et al.* 2010, Zepeda 2012, Day 2012). Secondly, PD should focus on implementing policy, improving teachers’ performance and focus on procedural and productive features of change. This is what Evans explains as ‘functional development’ (Evans 2008, p. 15). And thirdly, PD should focus on teachers’ ‘professional-related knowledge, understanding or reflective or

comprehensive capacity or competence' (Evans 2011, p. 867). Evans explains this as 'intellectual development' (Evans 2011, p. 867). In our view intellectual development balances attitudinal and functional development.

Our description of the purpose of PD therefore is: the combination of intellectual, attitudinal and functional development which provides teachers with knowledge, skills and professional understanding supporting teaching for the benefit of increasing student achievement, school professionalization and accompanying the implementation of educational reforms.

The questions that subsequently arise are when and why PD programmes are actually effective in terms of realizing our described purpose of PD. Answering these questions demands perspectives from the views of teachers' professionalism and system policy development. Furthermore it demands perspectives from both education in general and VET specifically.

Literature offers general insights in several core features of effective PD (Birman *et al.* 2000, Desimone *et al.* 2002, Desimone 2009). Although there is still limited evidence on the specific features of PD that enhance effectiveness (Wayne *et al.* 2008). Desimone (2009) elucidates to have found 'enough empirical evidence to suggest that there is in fact a consensus on a core set of critical features' (p. 183). She identified the features Content focus, Active learning, Collective participation, Duration and Coherence. Based on the work of Evans (2008, 2011), Desimone's findings, with the exception of the feature coherence, predominantly have an intellectual and motivational character and are primarily established from within the teaching profession. Our review utilizes a broader perspective and therefore enriches 'Desimone's findings' with the features: Onsite, Job embedded and, to a lesser degree Evidence and data driven (Villegas-Reimers 2003, Veen *et al.* 2010, Opfer and Pedder 2011, Zepeda 2012, Day 2012).

When considering pressures from outside the profession than PD has more functional characteristics. These characteristics are based on a growing knowledge intensity, demanding a higher procedural and productive focus from teachers (OECD 2005, Hargreaves and Shirley 2009, Day 2012). This means, although scarcely investigated, that PD core features are also based on delivering service and quality, a need for accountability so that learning outcomes and teacher performance can be indicated, measured and managed (Day 2012), and, finally, that PD core features stimulate the development of a performance culture (Sachs and Mockler 2012).

Considering the above it seems necessary to redefine existing core features of effective PD when confronted with the PD purposes presented by Evans (2008, 2011). This applies especially when we take into account that existing core features are predominantly focused on education in general and not specifically on VET.

For VET, which includes education and training, programmes designed for and typically leading to a particular job (OECD 2009), PD aims to produce 'useful outcomes in the labour market' (OECD 2009, p. 88). Therefore Desimone's core features (2011) lack consensus, concerning their effectiveness in VET PD. This is, amongst others, caused by the large number of stakeholders in VET as compared to primary and secondary education, like employers, social partners and unions (OECD 2009). Furthermore VET teachers are concerned with the challenges of 'double obsolescence' (Nieuwenhuis 2012, p.33). The former recognizes the shared responsibility between different stakeholders, the need for engagement with employers and unions to gain their support for policy as well as content implementation. The latter concerns substantial changes in teaching, simultaneous with

substantial in the professions VET teachers teach for. This needs attention in PD designs (OECD 2009).

Although lacking consensus on a core features framework for effective PD in VET several effectiveness-enhancing characteristics, for VET teacher PD, are described in the literature. First of all it is clear that VET PD needs to contribute to useful outcomes in the labour market and helps realizing specific legislated VET aims (OECD 2009, Glaude *et al.* 2011). Designing PD with social partners, unions and employers therefore seems necessary (Atwell 1999, Cort *et al.* 2004, Grollman 2008, Volmari *et al.* 2009, OECD 2009, Misra 2011). Furthermore VET PD includes subject and profession-specific knowledge and skills (McDaniel *et al.* 2010, Glaude *et al.* 2011, De Bruin 2012) and combining these with context-rich learning (Atwell 1999, Heikkinen 2002, Glaude *et al.* 2011). The combining of working and learning is also mentioned (Cedefop 2009, OECD 2009, Glaude *et al.* 2011). PD needs to be team-based, onsite and blended with at least informal learning characteristics (Cort *et al.* 2004, Volmari *et al.* 2009, Verbeek *et al.* 2009, Oudeman 2010, Nieuwenhuis 2012, De Bruin 2012). Moreover it needs a substantial focus on career paths and personal development for students (Cort *et al.* 2004, Volmari *et al.* 2009, Glaude *et al.* 2011). Evaluations of student performance through standardized frameworks of national assessment have demonstrated advantages in improved student performance, the value of qualifications promotes flexibility and innovations in learning (OECD 2009). Because of the often constructivism-based VET learning concepts, for VET students, more pedagogical knowledge and skills for teachers are assumed necessary (Atwell 1999, Cort *et al.* 2004, OECD 2009, Volmari *et al.* 2009, Oudeman 2010, De Bruin 2012). VET teachers need to utilize a real, adaptive and flexible combination of methods and need to develop more reflective areas of teacher professionalism (De Bruijn 2012). This is best established in an environment of collaboration and deliberation amongst professionals where collective, informal, learning processes contribute to the solving of professional problems and dilemmas (Verbiest and Timmermans 2008, Volmari *et al.* 2009, Verbeek *et al.* 2009, De Bruin 2012).

More recently the awareness is growing that ICT skills and methodologies are necessary in VET PD programmes in order to support the learning of students and the communication with stakeholders (Cort *et al.* 2004, Volmari *et al.* 2009, Cedefop 2009, Misra 2011). Additionally VET teachers need to develop competence in e-learning methods and techniques.

As in education in general VET also struggles with the incoherence in demands from inside and outside the profession. OECD (2009) indicates strong functional pressures by explaining that the development and implementation of (educational) policy depends on [...] ‘different stakeholders through strong institutions’ (p. 87). These findings were noted earlier by e.g. Heikkinen (2002) and Hodgkinson (2002). Furthermore Hodgkinson (2002), like Friedman and Phillips (2004) and Sachs and Mockler (2012) identifies audit and performance cultures in VET. This trend is tangible in the Dutch VET context (Coonen 2008, Oudeman 2010, Hooge 2013).

To summarise, like in education in general the presented feature of coherence (Desimone 2011) seems necessary to enrich PD in VET in order to increase its effectiveness. In European education teachers are generally seen as ‘actors who implement reforms or policies designed at higher administrative levels’ (Volmari *et al.* 2009, p. 50). More robust evidence is available from Australia. There, PD for VET teachers ‘has become an important policy lever in federal and state goals aimed at ratcheting up the quality of VET in publicly funded programmes’ (Cedefop 2009, p. 96). Oudeman (2010) establishes similarities for the Dutch context explaining that VET suffers from a constant dimness because of the lack of governmental direction. Therefore the VET institutes are overwhelmed with demanding

stakeholders and inconsistent policy interventions. ‘Because of the distance between executive boards, through (middle) management, and teachers there is no critical dialogue, demanding PD on all levels in VET institutions’ (Oudeman 2010, p. 19). Verbeek *et al.* (2009) stress that incoherent contexts can easily lead to transfer and motivational problems and actors falling back on traditional forms of PD. Therefore policies, management and executives need to align different and sometimes apparently opposed forms of learning and PD (Verbeek *et al.* 2009).

In order to create an overview of the characteristics of effective PD from both education in general and VET a redesigned framework on the core features of effective PD is necessary. We redesigned such a framework based on a review of research literature (Table 1). The various findings were categorized, grouped and labeled using keywords into 14 design requirements (DR). The purpose of the framework is to design effective PD in VET institutes.

| Table 1: DR for designing effective PD in VET and references | |
|--|---|
| 1 | Tolerance for complexity concerning changes in VET and the professions VET teachers teach. Heikkinen (2002), Mouwen (2008), OECD (2009), Verbeek <i>et al.</i> (2009), Sung (2010), Glaude <i>et al.</i> (2011), Opfer and Pedder (2011), Day (2012), Nieuwenhuis (2012) |
| 2 | Focus on aims of (VET) education. Glaude <i>et al.</i> (2011), Schoonhoven (2011), Ben-Peretz (2012) |
| 3 | Coherence. Garet <i>et al.</i> (2001), Heikkinen (2002), Hodgkinson (2002), Guskey (2002b), Borko (2004), OECD (2005), Desimone (2009), OECD (2009), Van Veen <i>et al.</i> (2010), Oudeman (2010), Desimone (2011), Opfer and Pedder (2011), De Bruin (2012), Day (2012), Sachs and Mockler (2012), Zepeda (2012) |
| 4 | Contribution to a professional identity. Lassnigg (2002), Evans (2008), McDaniel <i>et al.</i> (2010), Oudeman (2010), Ecorys (2011), Glaude <i>et al.</i> (2011), Huijts <i>et al.</i> (2011), Rijksoverheid (2011, 2012), De Bruin (2012), Dehing (2012), Nieuwenhuis (2012), BVMO (2013), MBO raad (2013), Onderwijscoöperatie (2013) |
| 5 | Collective, informal learning methodology. Flohr and Kamsma (2004), Verbeek <i>et al.</i> (2009), De Bruin (2012), Nieuwenhuis (2012) |
| 6 | Contribution to policy agendas as well as personal development. Day (2002; 2012), Villegas-Reimers (2003), Borko (2004), Friedman and Phillips (2004), OECD (2005), Evans (2008; 2009; 2011), Van Veen <i>et al.</i> (2010), Goodwin (2012), Zepeda (2012) |
| 7 | A focus on subject/ profession specific content, knowledge and skills. Atwell (1999), Heikkinen (2002), OECD (2009), Verbeek <i>et al.</i> (2009), McDaniel <i>et al.</i> (2010), Oudeman (2010), Van Veen <i>et al.</i> (2010), Desimone (2011), Opfer and Pedder (2011), Zepeda (2012), Glaude <i>et al.</i> (2011), De Bruin (2012) |
| 8 | Duration. Desimone (2011); Van Veen <i>et al.</i> (2010); Opfer and Pedder (2011); Zepeda (2012) |
| 9 | Active learning. Desimone (2011); Van Veen <i>et al.</i> (2010); Opfer and Pedder (2011); Zepeda (2012) |
| 10 | Team based and collective participation. Van Veen <i>et al.</i> (2010), Sung (2010), Desimone (2011), Glaude <i>et al.</i> (2011), Opfer and Pedder (2011), Nieuwenhuis (2012), Zepeda (2012) |
| 11 | Onsite and embedded in the job. Villegas-Reimers (2003), Verbeek <i>et al.</i> (2009), Van Veen <i>et al.</i> (2010), Opfer and Pedder (2011), Rubens <i>et al.</i> (2012) |
| 12 | Degree of evidence based and data driven. Van Veen <i>et al.</i> (2010), Zepeda (2012) |
| 13 | Relate to competence based VET focused on pedagogical and didactical knowledge and skills. Atwell (1999), Heikkinen (2002), OECD (2009), Verbeek <i>et al.</i> (2009), McDaniel <i>et al.</i> (2010), Oudeman (2010), Seezink and Poel (2010), De Bruin (2012), Glaude <i>et al.</i> (2011), Nieuwenhuis |

| | |
|----|------------------------------------|
| | (2012) |
| 14 | High quality assessment standards. |
| | OECD (2009) |

Once the purpose of PD is established and DR from literature is determined, the question arises how to evaluate the actual effectiveness of the DR in a specific PD activity.

Because of the complexity of VET PD it is essential that evaluation instruments used can cope with this complexity (Cervero 2000, Villegas-Reimers 2003, Borko 2004, Timperley 2008, Opfer and Pedder 2011, Day 2012). Opfer and Pedder (2011) explain this complexity as a combination of environment, school, teacher and PD activities that interacts in different combinations with varying intensities in specific contexts. Therefore PD has a situative perspective (Borko 2004).

Evaluation practice is most useful when it explores all these variables and their interrelationship (Day and Sachs 2004). Several frameworks suitable for coping with this complexity are presented in the literature (Holton 1996, Guskey 2000, Kirkpatrick and Kirkpatrick 2006, Coldwell and Simkiss 2011). We present a framework (Table 2) applicable to the demands of measuring PD effect on different levels of impact. This framework combines the work of Guskey (2000) and Kirkpatrick and Kirkpatrick (2006).

| Table 2: Levels of impact modified | |
|------------------------------------|--|
| Level of Impact | Description |
| Participants' attitudes | Participants' attitudes toward PD |
| Organizational support and change | Coherence of PD with e.g. policies, resources etc. |
| Participants' reactions | Primary opinions and feelings about PD |
| Participants' knowledge | PD contribution to learned knowledge |
| Participants' behaviour | PD contribution to applying learned in job |
| Outputs / outcomes | PD contribution to the occurrence of final results |

The impact model (Table 2) represents different levels of measuring PD impact and is applicable to the described environment, school, teacher and PD activities. The attitude and organizational support and change (OSC) levels seem necessary because PD effect is influenced by other factors than the PD programme. '(...)teacher change is also affected by individual and school factors' (Smith and Gillepsie 2007, p. 225). Individual aspects are teacher motivation (Day 2002, Kirkpatrick and Kirkpatrick 2006, Smith and Gillepsie 2007, Canrinus *et al.* 2011), teacher concerns (George *et al.* 2006, Smith and Gillepsie 2007, Hall and Hord, 2011) and teacher self-efficacy (Bandura 1997, Day 2002, Smith and Gillepsie 2007, Schwarzer and Hallum 2008, Canrinus *et al.* 2011, Thurlings *et al.* 2014). Following the attitude definition presented by Ajzen and Fishbein (2000), as the 'evaluation of an object, concept, or behaviour along a dimension of favor or disfavor, good or bad, like or dislike' (Ajzen and Fishbein 2000, p. 3), and the fact that modern theories of motivation are more focused on the relation of attitudes, beliefs, values, and goals with action (Eccles and Wigfield 2002) the constructs of motivation, concerns and self-efficacy emerge regularly.

School factors like organizational support, leadership, climate and/or supportive cultures are elements mentioned by Villegas-Reimers (2003), Kirkpatrick and Kirkpatrick (2006), Smith and Gillepsie (2007), Desimone (2009), Van Veen *et al.* (2010), Day (2012) and Evans (2014). In order to enhance PD efforts as much as possible it is necessary to implement the attitudes and OSC levels in our levels of impact framework. (Pre) measuring the attitude and OSC levels can help to clarify why PD programmes failed to be effective for participants. For the purpose of measuring the effectiveness of a mandatory PD programme, based on specific design requirements (Table 2), our framework seems adequate.

In our search to connect one or more PD design requirements to one or more levels of impact it is first necessary to establish a clear understanding of the levels of impact and their indicators of measurement. Table 3 describes the levels of impact and indicators chosen to measure effect. For each indicator references are presented.

| Table 3: Indicators for measuring impact | | |
|--|---|--|
| Level of Impact | Indicator for measurement | Reference |
| Participants' attitudes | General self-efficacy scale teacher | (Schwarzer and Jerusalem 1995, Schwarzer <i>et al.</i> 1999) |
| | Stages of concern | (Van den Berg and Vandenberghe 1981, George <i>et al.</i> 2008) |
| | Attitude towards profession | (Kwakman 1999) |
| Organizational support & change | Perceptions of received organizational support | (Guskey 2000) |
| Participants' reactions | Opinions and feelings about the programme | (Guskey 2000, Kirckpatrick and Kirckpatrick 2006) |
| Participants' knowledge | Changed attitudes, improved knowledge and/or skills | (Bloom <i>et al.</i> 1956, Guskey 2000, Anderson and Krathwohl 2001, Krathwohl 2002, Kirckpatrick and Kirckpatrick 2006, Marzano and Kendall 2007) |
| Participants' behaviour | Changed behaviour | (Hall and Hord 1987, Guskey 2000, Kirckpatrick and Kirckpatrick 2006, Hall <i>et al.</i> 2008) |

Research questions

Based on the aforementioned considerations this article aims at answering the following questions:

1. To what extent are the instruments used to measure effectiveness in different levels of impact of a specific redesigned PD programme valid and reliable?
2. To what extent is effectiveness of the specific PD programme related to specific design requirements (found in literature) used in the redesign of this programme?
3. To what extent are design requirements used in the specific PD programme robust, based on measured effectiveness on (one or more) levels of impact?
4. Which new design requirements and/or other consequences can be deduced from the evaluation of the effectiveness of specific design requirements in the redesigned PD programme?

Methodology

In order to measure effectiveness of PD programmes an assessor training programme, offered by a Dutch VET institute, was selected because of its mandatory character. In cooperation with the VET institute this programme was analysed on the presence of specific design requirements (DR). The analysis was conducted using the DR for effective PD framework presented earlier. This framework contains 14 DR which appear to enhance effectiveness in PD programmes in VET and its architecture is based on our description of the purposes of PD. Ten DR were designed in the programme. Furthermore, the programme consisted of six three- to four-hour meetings, was taught by experienced trainers and lead to a certification. In the programme self-tuition, guided by e-learning and varied with group meetings on site,

supported by coaching and intervision was designed and therefore had a blended character. Because the programme was offered infrequently our data collection was limited.

Sample

The characteristics of all participants involved are presented in Table 4. The sample consisted of 30 men (76,9%) and nine women (23,1%) which differs from the Dutch population of teachers in VET. In 2013, 47,9% men and 52,1% woman were employed in VET in The Netherlands (Stamos 2014). With respect to age the outcome is more comparable with VET as a whole (Stamos 2014). With regard to educational qualification this outcome is also roughly comparable to our sample. In 2009 70.5% of the teachers in VET had a Bachelor or Master degree, 18.5 % had a middle qualification added with a pedagogical training, 11% had no formal qualification (Stamos 2014). Furthermore teachers in our sample had a mean of 11.8 years of experience in VET and 66.7 % of the sample had assessor experience.

| Table 4: sample characteristics (n=39) | | | | |
|--|----------|------|-------|------|
| | | % | means | Sd |
| Gender | men | 76,9 | | |
| | women | 23,1 | | |
| Age (y) | | | 44 | 11.4 |
| Education | middle | 12,8 | | |
| | bachelor | 84,6 | | |
| | master | 2,6 | | |
| Experience in VET (y) | | | 11.8 | 8.92 |
| Experience as assessor (y) | | 66,7 | 2.1 | 3.45 |

Measures and procedures

In our study we used seven different instruments varying from quick reaction sheets, questionnaires, pen and paper test to observation sheets. Measurements were made on five levels of impact: attitude, reaction, learning, behaviour and organizational support and change. The outputs / outcomes level in our measurement framework (Table 2) was not investigated because of time limitations. The output level would concern inspection body satisfaction and ultimately useful outcomes in the future labour market. Data was collected through a mix of methods, in one or more components of the different levels of impact. Table 5 presents the different levels of impact, the indicators used, reference and methods. Per instrument an example of an item is presented.

| Table 5: Levels of impact, indicators and instruments | | | | | |
|---|-----------------------|---|--|---|--|
| level of impact | indicator | Reference (validity / reliability) | method | items | When |
| | Demographic data | | | 14 | Measurement before starting programme. |
| Attitude | Teacher Self Efficacy | (Swarzer <i>et al.</i> 1999, Schwarzer and Hallum 2008) | questionnaire Original instrument translated into Dutch. | 10 | |
| | | | | <i>When I try really hard, I am able to reach even the most difficult students.</i> | |
| | Stages of Concern | (George <i>et al.</i> 2008, Saunders 2012) | questionnaire Dutch | 35 | |

| | | | | | |
|------------------------|---------------------------------------|---|---|----|---|
| | | | instrument (Van den Berg and Vandenberghe 1981) was used. | | |
| | | <i>I am not concerned about the innovation at this time.</i> | | | |
| | Attitude & profession | (Kwakman 1999, 2001, 2003) | questionnaire | 12 | |
| | | <i>It is important that teachers respect school level made policy and agreements.</i> | | | |
| Organizational support | Perceptions of the received support | (Guskey 2000) | questionnaire | 18 | Measurement direct after completion |
| | | <i>My management facilitates this PD programme sufficiently.</i> | | | |
| Reaction | Reactions towards programme | Guskey 2000, Kirckpatrick and Kirckpatrick 2006) | self-developed quick reaction sheet | 7 | Measurement directly after each training meeting. |
| | | <i>What rating do you give today's meeting and why?</i> | | | |
| Learning | Knowledge of content of the programme | self-developed pen and paper test | | 8 | Measurement 2-4 weeks after completing. |
| | | <i>Because of this PD programme I now know and understand the assessment instruments I need to use.</i> | | | |
| Behaviour | Use of content of the programme | Self-developed observation-sheet | | 12 | Measurement 2-10 weeks after completing. |
| | | <i>Assessor explains assessment thoroughly to student.</i> | | | |

The choices for indicators, methods and instruments are based on those that have successfully been used in educational contexts and have had proper research concerning reliability, validity and robustness. This was the case for the attitude and OSC levels. For the reaction, learning and behaviour levels such instrument were not available. Therefore they were designed in a collaboration of researchers, PD trainers and trained assessors. These instruments were tested, evaluated and if necessary modified before being statistically analyzed.

In order to draw conclusions about the effectiveness of specific DR the following procedure was used. Each DR was operationalized using keywords based on literature. Then, results from each instrument, measuring one or more levels of impact, were compared with the keywords of each separate DR. The results of this comparison were used to relate DR to levels of impact and assess their functionality. Because of the complex, integral and situative character of the programme these comparisons did not fit well with the linear-rational levels of the impact model. This makes results interpretive.

Analysis

Of the 39 teachers who started the programme 34 completed the PD. Two teachers got a new job, two teachers dropped out because of illness and one teacher dropped out because of other PD obligations. The programme consisted of six meetings which were attended 4.5 times on average. Several participating teachers did not complete all measures which explains the different sample sizes in different measures.

Data were analysed using IBM SPSS 21; the reaction level of impact was analysed qualitatively. . Limitations for reliable results is our sample size which can cause sampling inadequacy. The Kaiser-Meyer-Olkin test was used to verify the sampling adequacy. The internal consistency of the subscales was tested using the Cronbach's α test.

In order to measure knowledge and specific behaviour, the learning and behaviour levels of impact were analysed using frequencies in the descriptive statistics.

The reaction level of impact, which consisted largely of open ended questions was analysed qualitatively using a specific coding scheme where feedback of respondents was coded and categorized in meaningful categories. These categories were related to the DR used in the PD programme. Grades were given by participants, per meeting, for trainer and meeting satisfaction. These data were analysed using frequencies in the descriptive statistics mode.

Results

For the attitude and organizational support levels of impact the Kaiser-Meyer-Olkin test verified the sampling adequacy, valuing all variables above the acceptable .5 limit (Field 2013). The internal consistency of all subscales overall showed acceptable Cronbach's α (George and Mallery 2003) except the stage 4 of SoC subscale, which yielded internal consistency of .54 indicating a weaker internal reliability. Table 6 presents the levels of impact, instruments used, number of items, KMO values and Cronbach's α . Considering the results it can be concluded that our instruments are of sufficient quality with the exception of stage 4.

| Table 6: Instrument quality analyses | | | | | |
|--------------------------------------|--|---------|-------|-------|---------------------|
| Level of impact | Instrument | | items | K-M-O | Cronbach's α |
| Attitude | Teacher Self Efficacy (n=39) | | 10 | .71 | .83 |
| | Stages of Concern (n=39) | stage 0 | 5 | .73 | .77 |
| | | stage 1 | 5 | .61 | .63 |
| | | stage 2 | 5 | .66 | .63 |
| | | stage 3 | 5 | .64 | .74 |
| | | stage 4 | 5 | .58 | .54 |
| | | stage 5 | 5 | .67 | .82 |
| | | stage 6 | 5 | .57 | .62 |
| | Attitude & profession (n=32) | | 12 | .64 | .75 |
| Organizational support | Perceptions of the received support (n=32) | | 17 | .53 | .88 |

The results on the attitude measures are presented in table 7. Where it concerns teacher self-efficacy (TSE), the level of perceived TSE of participants is high. All items related to job-accomplishment, skill development on the job, social interaction with students, parents, and colleagues, and coping with job stress (Schwarzer and Hallum 2008) show high scores.

Regarding the stages of concern (SoC) the peak concerns were stage 4 (consequence) and stage 5 (collaboration). This suggests that firstly, participants' concerns focus on the programme's impact on students; its relevance for students, student outcomes and the changes needed to influence student outcome. Secondly, concerns focus on coordination and cooperation with others regarding the use of the PD programme (Hall and Hord 2011). Stage 0 shows very low scores indicating substantial concern about the programme.

Finally the attitude and profession (A&P) measure showed high scores. Here it concerns professional attitudes, with regard to the realization of educational aims, improvement of personal performance and shared responsibility for the school organization (Kwakman 1999). Our participants agree with the propositions presented in the instrument, in particular the realization of educational aims. Limited but prominent doubts concern the shared responsibility for the school organization. In the face of all scores our population

clarifies professional attitudes especially as the realization of educational aims and improvement of personal performance. Performing other, organizational tasks is not automatically considered as part of one's professionalism.

| Table 7. Attitude level of impact, descriptive statistics (n=39) | | | | | |
|--|---|--------|--------|------|------|
| measu re | Scale aspects | | scores | | |
| | | | range | mean | sd |
| TSE | Job accomplishment, skill development on the job, social interaction with students, parents, and colleagues and coping with job stress. | | 1-4 | 3.49 | .55 |
| SoC | 0 Unconcerned | Self | 0-7 | 1.71 | 1.48 |
| | 1 Informational | | | 3.95 | 1.74 |
| | 2 Personal | | | 4.17 | 1.84 |
| | 3 Management | Task | | 2.39 | 1.70 |
| | 4 Consequence | Impact | | 4.29 | 1.64 |
| | 5 Collaboration | | | 4.91 | 1.60 |
| | 6 Refocusing | | | 3.15 | 1.70 |
| A&P | Realizing educational aims, improvement personal performance and shared responsibility for the school organization. | | 1-5 | 4.23 | .69 |

The results from the organizational support and change measure are presented in Table 8. Especially the aspects of organization policies, recourses and management leadership and support score high while provision of time is less appreciated.

| Table 8. Organizational support and change level of impact, descriptive statistics (n=32) | | | | |
|---|-----------------------------------|--------|------|------|
| measure | Scale aspects | scores | | |
| | | range | mean | sd |
| OSC | Organization policies | 1-7 | 5.81 | 1.17 |
| | Resources | | 5.23 | 1.16 |
| | Protection from intrusion | | 4.95 | 2.03 |
| | Openness to experimentation | | 4.77 | 1.52 |
| | Collegial support | | 4.45 | 1.81 |
| | Management leadership and support | | 5.13 | 1.84 |
| | Recognition of success | | 4.50 | 1.25 |
| | Provision of time | | 3.91 | 1.71 |
| | Overall | | 4.53 | 1.50 |

The results from a pen and paper test on the learning level of impact are presented in Table 9. Per open question, eight in total, a score of 10 points could be awarded (weight). Sufficient gain in knowledge was registered when respondents scored 50 points or more (caesura). From the eight questions asked only two (2 and 5) show high scores, two questions show average scores (3 and 7), all other questions (1, 4, 6 and 8) score low.

Results show that only two out of 22 respondents who completed the tests, show sufficient scores, scoring 50 points. Nine participants scored between 40 and 50 points, all others (11) scored less than 40 points. These results in knowledge are very limited. Questions two and five were remarkably well answered. A plausible explanation is that question two and five consist of knowledge this population needs and uses in more activities than just assessing and is not completely new knowledge.

| Table 9: Learning level of impact, descriptive statistics (n=22) | | | |
|--|-------|------|----|
| Questions | range | mean | sd |

| | | | | |
|----|--|------|------|------|
| 1. | Describe minimum three demands, according to the national qualification documents, set on the profession students are educated for. | 0-10 | .68 | 1.59 |
| 2. | Describe four of the six primary tasks described in the specific national qualification documents. | | 8.41 | 3.16 |
| 3. | Describe the separate parts of the assessment and explain the relationship between these parts. | | 4.23 | 2.51 |
| 4. | Describe the assessor tasks sequenced. | | 3.77 | 2.65 |
| 5. | Which method does an assessor use during the criterion based interview and what do the separate parts of the method imply? | | 9.09 | 2.94 |
| 6. | In what specific way and why do the assessment instruments help the assessor during the assessment? | | 2.41 | 1.68 |
| 7. | Because of the assessor training I know and understand the concepts validity, reliability and transparency. Describe the essence of each separate concept. | | 6.18 | 3.50 |
| 8. | Describe when you, as an assessor, judge an assessment as entirely sufficient. | | 2.64 | 2.08 |

Results on the behaviour level of impact are presented in table 10. A score of 1 represents that the behaviour could not be observed at all, 2 in case behaviour was observed partly and 3 behaviour was fully observed. In order to establish behavioural results from the PD programme a score ≥ 2 was determined necessary. Overall conclusion is that the trained behaviour is well executed. Striking however are the low scores for behaviours 7, 9, 11 and 13. These behaviours are very similar and comparable in the procedure where teachers have to summarize findings as part of the method used. It seems plausible that summarizing was not trained sufficiently in the programme.

| Table 10: Behaviour level of impact, descriptive statistics (n=24) | | | | |
|--|---|-------|------|-----|
| Observed Behaviour | | range | mean | sd |
| 1 | Assessor discusses assessment protocol with colleague assessor. | 1-3 | 2.58 | .88 |
| 2 | Assessor ensures an appropriate environment during the interview. | | 3.00 | .00 |
| 3 | Assessor creates an atmosphere in which the candidate is at ease. | | 2.88 | .32 |
| 4 | Assessor explains the complete assessment to the student. | | 2.33 | .82 |
| 5 | Assessor uses the STAR instrument described in the protocol. | | 2.46 | .72 |
| 6 | Assessor reports the situation literally and thoroughly. | | 2.38 | .65 |
| 7 | Assessor summarizes findings from the situation. | | 2.00 | .93 |
| 8 | Assessor reports the task literally and thoroughly. | | 2.29 | .68 |
| 9 | Assessor summarizes findings from the task. | | 1.71 | .86 |
| 10 | Assessor reports action literally and thoroughly. | | 2.46 | .66 |
| 11 | Assessor summarizes findings action. | | 1.83 | .82 |
| 12 | Assessor reports results literally and thoroughly. | | 2.38 | .71 |
| 13 | Assessor summarizes findings results. | | 2.00 | .93 |
| 14 | Assessor fills in the protocol independent from colleague assessor. | | 2.17 | .76 |
| 15 | Assessor consults with colleague assessor to present one outcome. | | 2.54 | .98 |
| 16 | Assessor determines the (concept) result of the assessment. | | 3.00 | .00 |

The reaction level of impact analysis resulted in the overview presented in table 11. It shows the questions asked and the meaningful categories based on similarities and patterns in respondents' most frequent answers. The total amount of answers is presented in the last column.

Other categories were detected too, but with less convincing frequencies. For example “more discussion”, “more specific practice” and “prepare specifically”.

Explicit feedback from teachers about the programme concerns the realistic practice, using real live students combined with moments when teachers work together share feedback and reflect. Specific suggestions for improvement were made when appreciated elements were limited during a training meeting. It concerns realistic training, sufficient dialogue and according trainer behaviour. Prominent are suggestions considering time in the category time management. The duration appears to be sufficient but the positioning of the programme, time of day and place in a week are important. Preferably earlier in the day (morning) and early in the week (Mondays, Tuesdays and Wednesdays). A final conclusion is the awareness of teachers of their limited efforts to prepare and study. This was hardly done but in most cases with good intentions for the next meeting.

The training was evaluated quantitatively using marks (1-10) at the end of each meeting. The average score of all five meetings is 6.12 with a standard deviation of 1.36. The trainer was graded a 7.10 average over five meetings with a standard deviation of 1.19. This concludes an overall sufficient satisfaction with the meetings and trainer.

| Table 11: Reaction level of impact → Frequencies meaningful units (n=39) | | | | |
|--|---|-----------------------------|-----------|------------------|
| Questions asked | | Meaningful category | Frequency | Total statements |
| 1 | What were you satisfied with in this meeting, what went well? | 1.1 Realistic practice | 37 | 98 |
| | | 1.2 Working together | 19 | |
| 2 | What are your suggestions for improvement? | 2.1 Trainer behaviour | 19 | 65 |
| | | 2.2 Time management | 14 | |
| 3 | Articulate one wish for the next meeting? | 3.1 More realistic practice | 24 | 69 |
| 4 | What are you going to do to contribute optimally during the next meeting? | 4.1 Prepare better | 36 | 64 |

In table 12 we have summarized which specific design requirements showed effects in the programme when measured on specific levels of impact. The DR present in the programme were related to all the measurements (table 5) using keywords. Positive effects are presented with +; negative effects with – and no effect with 0.

It was expected that, because of the mandatory character of the programme, positive attitudes and professionalism of participants would reinforce specific DR implemented in the programme. In the attitude LOI we have seen that our population perceives high rates of self-efficacy, strong professional attitudes and is focused on student impact. Participants were motivated and optimistic to start this programme. This is reflected in the amount of DR that according to participants frequently contributed to effectiveness.

Based on our measuring framework the DRs that contribute most frequently to the effectiveness of the programme are “tolerance for complex contexts”, “contribution to a professional identity” and “team based and collective participation”. Remarkable is the duration DR because of negative but explicit feedback. Facilitated time is appreciated most, early in the week and early on the day. The OSC and reaction levels of impact related most often to DRs. The learning level of impact hardly measured any result.

| Table 12: Effectiveness related to specific design requirements and levels of impact | | | | | |
|--|------------------|-----|-----|-----|-----|
| Design requirements implemented in PD programme | levels of impact | | | | |
| | ATT | OSC | REA | LEA | BEH |

| | | | | | |
|--|---|---|---|---|---|
| Tolerance for complex contexts concerning changes in the VET profession AND the professions VET teachers teach | + | + | + | + | + |
| Coherence | + | + | 0 | 0 | + |
| Contribution to a professional identity | + | + | + | 0 | + |
| Collective, informal learning methodology | 0 | + | + | 0 | + |
| Contribution to policy agendas as well as personal development | + | + | 0 | 0 | + |
| A focus on subject/ profession specific content, knowledge and skills | 0 | 0 | + | 0 | 0 |
| Duration | 0 | - | - | 0 | 0 |
| Active learning | + | + | + | 0 | 0 |
| Team based and collective participation | + | + | + | 0 | + |
| On site and embedded in job | 0 | + | + | 0 | + |

As a final result of our study means, standard deviations, reliability coefficients (diagonal) and Spearman correlations are presented in Table 13.

In our analysis we focus on the main constructs namely (the composed variables) attitude, OSC, learning, behaviour, and reaction. The attitude construct includes the teacher's self-efficacy (TSE), the professional attitudes (PA) and the stages of concerns (SOC) variables. The reaction construct includes meeting reactions (REA M) and trainer reactions REA T) variables. All other variables, namely organizational support and change (OSC), learning (LEAR) and behaviour (BEH) are independent constructs.

The strongest and most meaningful correlations were found between: teacher self-efficacy (TSE) and professional attitude ($r=.47$, $p<.01$); teacher self-efficacy (TSE) and SOC 6 ($r=.42$, $p<.01$); professional attitude (PA) and SOC 5 ($r=.50$, $p<.01$); SOC 0 (SOC 0) and perceived organizational support ($r=-.55$, $p<.01$); SOC 3 (SOC 3) and perceived organizational support ($r=-.42$, $p<.01$); SOC 4 (SOC 4) and behaviour ($r=.46$, $p<.01$) and meeting reactions (REA M) and reaction trainer ($r=.69$, $p<.01$).

| Table13: means, standard deviation, reliability coefficients (diagonal) and correlations | | | | | | | | | | | | | | | | | |
|--|------|------|-------|----------|-------|--------|-------|------|--------|-------|-------|------|------|------|------|----------|-----|
| | M | SD | range | Attitude | | | | | | | | | 10. | 11. | 12. | Reaction | |
| | | | | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | | | | 13. | 14. |
| 1. TSE (n39) | 3.49 | .55 | 1-4 | .83 | | | | | | | | | | | | | |
| 2. PA (n39) | 4.23 | .69 | 1-5 | .47** | .75 | | | | | | | | | | | | |
| 3. SOC0 (n39) | 1.71 | 1.48 | 0-7 | -.37* | -.25 | .77 | | | | | | | | | | | |
| 4. SOC1 | 3.95 | 1.74 | 0-7 | .13 | .20 | .07 | .63 | | | | | | | | | | |
| 5. SOC2 | 4.17 | 1.84 | 0-7 | .11 | .23 | -.08 | .73** | .63 | | | | | | | | | |
| 6. SOC3 | 2.39 | 1.70 | 0-7 | -.34* | -.15 | .66** | .38** | .27 | .74 | | | | | | | | |
| 7. SOC4 | 4.29 | 1.64 | 0-7 | -.00 | .25 | -.01 | .48** | .35* | .22 | .54 | | | | | | | |
| 8. SOC5 | 4.91 | 1.60 | 0-7 | .32* | .50** | -.38** | .21 | .15 | -.32* | .37* | .82 | | | | | | |
| 9. SOC6 | 3.15 | 1.70 | 0-7 | .42** | .27 | .03 | .44** | .18 | .23 | .49** | .45** | .62 | | | | | |
| 10. OSC (n32) | 4.81 | 1.55 | 1-7 | .27 | .37* | -.55** | .13 | .27 | -.42** | -.01 | .22 | -.14 | .88 | | | | |
| 11. LEAR (n22) | 4.28 | 2.63 | 0-10 | -.14 | .14 | -.24 | .00 | -.03 | .08 | .26 | .36 | .14 | .16 | - | | | |
| 12. BEH (n24) | 2.25 | .67 | 1-3 | .14 | .02 | -.01 | .08 | .21 | -.06 | .46* | .29 | .09 | .00 | .15 | - | | |
| 13. REA M (n30) | 6.12 | 1.36 | 1-10 | .01 | -.07 | .14 | .11 | .27 | .20 | -.13 | -.31 | -.18 | .37* | -.03 | -.20 | - | |
| 14. REA T (n30) | 7.10 | 1.19 | 1-10 | .27 | .04 | .16 | .02 | .17 | -.22 | -.15 | -.07 | -.05 | .29 | .14 | -.19 | .69** | |
| **. Correlation is significant at the 0.01 level (1-tailed). | | | | | | | | | | | | | | | | | |
| *. Correlation is significant at the 0.05 level (1-tailed). | | | | | | | | | | | | | | | | | |

Within the attitude construct self-efficacy and an optimistic professional attitude are associated with impact concerns. A modest correlation was found between the professional attitude construct and organizational support and change. In the stages of concern in the attitude construct negative correlations were found with organizational support and change, especially in the self- and task concerns. These stages consider time and personal management items. Stage 4 which reflects student impact concerns correlates negatively with the behaviour construct. Finally the reactions per meeting are associated with participants'

reactions towards the trainer. An interesting result is stage 1 (SOC) which relates to many of the other stages: 2, 3, 4 and 6.

We expected that effectiveness of the programme could be explained because of specific DRs implemented in the programme. Furthermore we expected that feedback from the attitude and organizational support and change levels could explain the functionality of specific DRs.

Although scores in the attitude and organizational levels of impact were high, the lack of effect in the learning level of impact could be explained because of the perceived shortage of provision of time. This also seems to be the case with respect to the reaction level of impact. Here participants' feedback suggested the programme should be earlier in the day and earlier in the week.

Conclusions and discussion

Considering our first research question concerning the validity and reliability of instruments, we have provided statistical evidence that the quality of the instruments used in the attitude and organizational support and change levels of impact was adequate (Table 6). The instruments used for the other levels of impact were self-constructed, in a collaboration of researchers, PD trainers and trained assessors and specifically designed, validated and tested, with good results, for a specific PD programme.

To answer research questions two and three concerning the effectiveness of the programme related to specific design requirements and specific levels of impact, we will use the framework presented by Nieveen (2010) describing four generic criteria to draw conclusions for high quality interventions in design based research: relevance, consistency, practicality and effectiveness.

Firstly, was the PD intervention relevant? In our study the PD intervention was demanded by the inspection body as an accountability policy mechanism and from that point of view relevant for the school. Furthermore it was also an intervention at the core of teachers' work; assessing students' achievements. The relevance, based on our research into PD effectiveness, was established because of the scientific backbone of our design requirements and methodology, the triangulation of data source, data type, data analysis, method and theory and the collaboration of researchers, PD trainers, teachers and management.

Secondly, was the PD intervention consistent? Construct validity, as described in Nieveen's (2010) framework, was established by detailed analyses of theory and the use of operational definitions. Furthermore the collaboration of PD researchers, PD trainers, participating teachers in PD practice and school management also contributed to construct validity. However, because of mono operation bias (knowledge base 2015) the construct validity, for this point in time needs to be labelled as work in progress. The programme only used a single version, in a single place at a single point in time. The full breadth of the programme, therefore, cannot be captured. Furthermore a threat to the construct validity is the evaluation apprehension, concerning the human tendency to perform well in measurements (knowledge base 2015).

Thirdly, practicality. This was confirmed by participants during and after the programme as our results show (Table 11). The programme can be used in different contexts, with different populations.

Finally, effectiveness is determined by whether the programme helped solving a complex educational problem, putting forward the practical relevance and labelled as being use-inspired, applied oriented and/or socially responsible research (Plomp and Nieveen 2010).

Considering our findings we can conclude that the programme was only partially effective.

Based on Evans' (2008, 2011) framework, attitudinal and intellectual development was only partly achieved because of the lack of learning and professional related knowledge, the programme was not effective at the learning level of impact (table 9). Furthermore impact and outcomes of the programme, as inspection body satisfaction, could not be measured. More effectiveness concerned the functional development (Evans 2008, 2011) considering the implementation of mandatory policy and results in the behaviour level of impact. These were sufficient (table 10). This concludes a sufficient use or gain in intellectual capacities, positive attitudes and student impact concerns (Table 7).

The effectiveness in terms of the design approach describes the functioning of our programme and theoretical and empirical evidence (or strong indications) of the DR, procedures and guidelines.

The functioning of the PD programme with its essence in the use of specific DRs and when and how these DRs work are summed up in table 12. Especially for the DR: "tolerance for complex contexts", "contribution to a professional identity" and "team-based and collective participation" we have interpreted indications for effectiveness. These DRs were measured mainly in the OSC and reaction levels of impact. Considering the establishment of procedures and guidelines, our research shows these can be indicated as solid and robust with consistency marked as work in progress.

Finally, research question four which focused on lessons learned concerning the use of DR explicitly gives feedback on the DR of duration. Teachers respond to this DR, assuming an increase in effectiveness of the programme, when time is explicitly offered and the programme is offered early in the week and early in the day. Furthermore the DR: focus on subject/profession specific content, knowledge and skills proved irrelevant this specific assessor-type of training.

Discussion

In this study we aimed to provide explanatory insights in the how and why of the functioning of our redesigned PD programme. Following a design approach we provided a scientific architecture for the programme. The outcomes of our study indicate a successful, formatively evaluated prototype PD programme. However our methodical architecture needs discussion. Although the models of both Kirkpatrick and Guskey were enormously influential the models have not gone unchallenged (Alliger and Janak 1994, Holton 1996, Coldwell and Simkins 2011). The causality assumptions that a positive effect in one level is necessary to have effect in the next are disputable. Similar criticisms applies for Guskey's (2000) model. Additional shortcomings of the models when measuring the effectiveness of PD programmes are described by Holton (1996). In particular when an effect is not found to occur and many intervening factors remain unmeasured conclusions about defaults in the PD programme may not be valid. Because of the compulsory character of the programme this seems quite plausible in this study. Examples are the negative correlations in the OSC construct, concerning time aspects and limitations in learning effect. Low scores can be explained as participants themselves are aware of their limited efforts to prepare and study amongst others because of the compulsory character of the programme and their perceived full and busy agendas. Explanations can also be attributed to limited regulated professional standards in Dutch VET (Mc Daniel *et al.* 2009), lacking extensive on-the-job training and self-regulated learning in the organizational culture which limits self-tuition. Cedefop amongst others also reports that VET in Europe has a weak PD culture (Cort *et al.* 2004, Cedefop 2009). An

explanations might also be found in the more practical orientation and behaviour of VET teachers who preferably learn by doing. McDaniel, *et al.* (2010) mention that the concept of teacher professionalism with high levels of self-regulated learning within institutions in The Netherlands has not yet developed very well.

The mentioned full and busy agendas of teachers, preventing them from studying is elaborated by Volmari *et al.* (2009): ‘particular challenges, in the work of VET teachers, are the increase in administrative tasks and responsibilities’ (2009, p. 50). These tasks however could, assumingly, correspond with our reported high levels of professional attitudes and self-efficacy, so they should not automatically prevent professional teachers from studying. The compulsory character of the programme most likely limits teachers to prepare sufficiently. Finnigan and Gross (2007) for example, who researched the influence of accountability policy on teacher motivation, reports that teachers respond less to incentives or threats built into specific policy like our PD programme. In our case this could explain the lack of learning. It explains that teachers do not perceive the programme as theirs which limits internalization and attitudinal and intellectual development. It does not mean that teachers are not concerned about the programme as we have seen in the stages of concern measure (Table 7). The very low score in stage 0 suggest shows that most teacher are concerned about the change initiative. In terms this means even more focus on the attitudinal and OSC levels when preparing for these types of PD.

The outcomes of our study, presented in tables 12 and 13, provide sufficient insights to continue the use of key characteristics, methods and procedures in the (re)design of PD programmes. Yet the specific DR used in the PD programme need discussion concerning robustness and generalization. Some of the DR depend on trainer behaviour and are therefore vulnerable for fluctuations.

Furthermore our DR were not developed in cooperation with social partners, unions, participants and employers making them perhaps incomplete. Finally the effectiveness of specific DR in the programme is interpretative because of the complex, integral and situative character of the programme. This characteristic did not fit well with the linear-rational levels of the impact models of Guskey (2000) and Kirkpatrick and Kirkpatrick (2006).

Where it concerns the scientific body of knowledge we suggest to specify Desimone’s ‘core set of critical features’ (2009, p. 183). Perhaps critical findings should be specified to specific types of education because, as we have seen, VET has some specific demands in comparison to education in general such as useful outcomes in the labour market. Furthermore distinction can be made between attitudinal and functional based PD (Evans, 2008; 2011). Both need specific PD DRs. For example the “focus on subject/ profession specific content, knowledge and skills” DR which in our study was not relevant. This is prominent because literature often attributes PD effect because of this characteristic (Van Veen *et al.* 2010, Desimone 2011, Opfer and Pedder 2011, Zepeda, 2012). An explanation might be that this specific programme was specifically designed for assessment purpose. Profession and/or content specifics are not necessarily relevant in these types of programmes.

Limitations and recommendations for further research

First of all further research is necessary because of the character of our approach of design based research. Findings should be evaluated in several prototype cycles to confirm key characteristics and methods and procedures used in order to increase construct validity for generalization purpose (Plomp and Nieveen 2010, knowledge base 2015). Striving toward context free generalizations however will be challenging because of the usual holistic

approach of design research with a lack of emphasizing isolated variables (Van den Akker *et al.* (2006 cited Plomp and Nieveen 2010)).

Secondly, our conclusions might be biased because feedback from the attitudinal, organizational support and reaction levels reflect participants' perceptions and therefore are indirect measures and not direct evidence. This is especially the case when sample sizes are limited which is our third relativizing remark. Although data was collected triangulated the small sample size limits validity and generalization. Furthermore it could also explain the .54 Cronbach's α in stage 4 of our stage of concern measure. This is interesting because George *et al.* (2006) presented internal reliability coefficients for all stages from at least seven major studies never showing Cronbach's α below .71 in stage four. Our suggestion for future research is the extension of items in the stage 4 questionnaire.

Recommendations for further research concern validation and/or clarification of our findings, especially our intervention theory and design principles. Construct validity can be enhanced by researching these principles in multi contexts, in multi PD programmes and, in multi points in time.

Specific attention should be given to the self-guided cognitive learning ability of teachers as it seems to limit effectiveness because of the mandatory character of the programme. Research on this aspect is necessary because the character of educational systems demands mandatory and therefore functional professional development. Research on DR in order to enhance the effectiveness of PD programmes probably is just one approach in reducing incoherence between the goals of education, teachers PD, and the 'politics of accountability' (Ben-Peretz 2012, p.57). More cross-over educational, psychological, sociological, policy and organizational research is necessary to enhance the quality of PD.

References

- Ajzen, I., and Fishbein, M., 2000. Attitudes and the Attitude-Behaviour Relation: Reasoned and Automatic Processes. *European Review Of Social Psychology*, 11 (1), 1-33.
- Alliger, G., and Janak, E., 1994. Kirkpatrick's levels of training criteria: thirty years later. In: C. Scheier, C. Russell, R. Beatty and C. Baird, eds. *The training and development sourcebook*. Amherst, MA: HRD Press, 219–228.
- Anderson, L.W., and Krathwohl, D.R., ed., 2001. *A taxonomy for learning, teaching and assessing, a revision of Bloom's taxonomy of educational objectives*. New York: Pearson.
- Attwell, G., 1999. New roles for vocational education and training teachers and trainers in Europe: a new framework for their education. *Industrial And Commercial Training*, 31 (5), 190-200.
- Bandura, A., 1997. *Self-efficacy, The exercise of control*. USA: W.H. Freeman and Company.
- Béduwé, C., et al., 2009. New and emerging issues in vocational education and training research beyond 2010. In: CEDEFOP. *Modernising vocational education and training. Fourth report on vocational education and training research in Europe*. Luxembourg: Office for Official Publications of the European Communities, 17–72.
- Ben-Peretz, M., 2012. Accountability vs. teacher autonomy: an issue of balance. In: C. Day, ed. *The routledge international handbook of teacher and school development*. New York: Routledge, 57-66.
- Birman, B.F., et al., 2000. Designing Professional Development That Works. *Educational Leadership*, 57 (8), 28-33.
- Bloom, B.S., et al., 1956. *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay Company.
- Borko, H., 2004. Professional development and teacher learning-mapping the terrain. *Educational researcher*, 32 (3), 3-15.

- Bruijn, de E., 2012. Teaching in Innovative Vocational Education in the Netherlands. *Teachers And Teaching: Theory And Practice*, 18 (6), 637-653.
- Canrinus, E.T., *et al.*, 2011. Profiling Teachers' Sense of Professional Identity. *Educational Studies*, 37 (5), 593-608.
- Cervero, R.M. (2000). Trends and issues in continuing professional education. In: V.W. Mott and B.J. Daley, eds. *Charting a course for continuing professional education: Reframing professional practice. New Directions for Adult and Continuing Education*, no.86. San Francisco: Jossey-Bass, 3–12.
- Coldwell, M., and Simkins, T., 2011. Level models of continuing professional development evaluation: a grounded review and critique. *Professional Development In Education*, 37 (1), 143-157.
- Coonen, H., 2008. Professioneel leraarschap en onderwijsinnovatie. *Tijdschrift voor Hoger Onderwijs en Management*, 3, 29-34.
- Cort, P., Harkonen, A., and Volmari, K., 2004. *PROFF—Professionalization of VET teachers for the future*. Luxembourg: Office for the Official Publications for the European Communities.
- Day, C., 2002. School Reform and Transitions in Teacher Professionalism and Identity. *International Journal Of Educational Research*, 37 (8), 677-692.
- Day, C., ed., 2012. *The routledge international handbook of teacher and school development*. New York: Routledge
- Day, C., and Sachs, J., eds., 2004. *International handbook on the continuing professional development of teachers*. Berkshire: Open university press.
- Desimone, L.M., 2009. Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educational Researcher*, 38 (3), 181-199.
- Desimone, L.M., 2011. A primer on effective professional development. *kappanmagazine*, (summer issue), 1, 28-31.
- Desimone, L., *et al.*, 2002. Effects of Professional Development on Teachers' Instruction: Results from a Three-year Longitudinal Study. *Educational Evaluation And Policy Analysis*, 24 (2), 81-112.
- European Centre for the Development of Vocational Training (CEDEFOP), 2009. *Modernising vocational education and training. Fourth report on vocational education and training research in Europe: Synthesis report*. Luxembourg: Publications Of Ece of the European Union.
- Evans, L., 2008 .Professionalism professionalism and the development of educational professionals. *Britisch journal of education studies*, 1 (56), 20-38.
- Evans, L., 2011. The ‘shape’ of teacher professionalism in England: professional standards, performance management, professional development and the changes proposed in the 2010 White Paper. *British Educational Research Journal*, 37 (5), 851-870.
- Evans, L., 2014. Leadership for professional development and learning: Enhancing our understanding of how teachers develop. *Cambridge Journal Of Education*, 44 (2), 179-198.
- Eccles, J.S., and Wigfield, A., 2002. Motivational beliefs, values, and goals. *Annual Review Of Psychology*, 53 (1), 109-132.
- Field, A., 2013. *Discovering statistics using IBM SPSS statistics*. 4th ed. London: Sage.
- Finnigan, K.S., and Gross, B., 2007. Do accountability policy sanctions influence teacher motivation? Lessons from Chicago’s low-performing schools. *American Educational Research Journal*, 44, 594–629.

- Friedman, A., and Phillips, M., 2004. Continuing Professional Development: developing a vision. *Journal of education and work*, 3 (17), 361-276.
- George, A.A., Hall, G.E., and Stiegelbauer, S.M., 2006. *Measuring implementation in schools: The Stages of Concern Questionnaire*. Austin, TX: SEDL
- George, D., and Mallery, P., 2003. *SPSS for Windows step by step: A simple guide and reference. 11.0 update*. 4th ed. Boston: Allyn & Bacon.
- Glaude, M., et al., 2011. *Pedagogisch-didactisch handelen van docenten in het middelbaar beroepsonderwijs. Literatuurstudie*. s-Hertogenbosch-Utrecht: ECBO.
- Grollmann, P., 2008. The Quality of Vocational Teachers: Teacher Education, Institutional Roles and Professional Reality. *European Educational Research Journal*, 7 (4), 535-547.
- Guskey, T.R., 2000. *Evaluating Professional Development*. California: Corwin Press Inc.
- Guskey, T., 2002. Professional Development and Teacher Change. *Teachers And Teaching Theory And Practice*, 8 (3), 381-391.
- Hall, G.E., Dirksen, D.J., and George, A.A., 2008. *Measuring implementation in schools: Levels of Use*. Austin, TX: SEDL.
- Hall, G.E., and Hord, S.M., 1987. *Change in schools, Facilitating the Process*. Albany: State University of New York Press.
- Hall, G.E., and Hord, S.M., 2011. *Implementing change: Patterns, principles and Potholes*. 3rd ed. Upper Saddle River: Pearson.
- Hargreaves, A., and Shirley, D., 2009. *The fourth way, the inspiring future for educational change*. California: Thousand oaks.
- Heikinnen, A., 2002. Transforming VET policies and professionalism: a view from Finland. In: Nijhof, W.J., Heikinnen, A. and Nieuwenhuis, L.F.M., eds. *Shaping flexibility in Vocational Education and Training*. Dordrecht: Kluwer Academic Publishers, 207-225.
- Hodgkinson, P., 2002. The practice of a new VET profession. In: Nijhof, W.J., Heikinnen, A., and Nieuwenhuis, L.F.M., eds. *Shaping flexibility in Vocational Education and Training*. Dordrecht: Kluwer Academic Publishers, 259-268.
- Holton, E., 1996. The flawed four-level evaluation model. *Human resource development quarterly*, 7 (1), 5-22.
- Hooge, E., 2011. *Besturing van Autonomie*, Tilburg: Tilburg University.
- Kirkpatrick, D.L., and Kirkpatrick J.D., 2006. *Evaluating Training Programmes*. 3rd ed. San Francisco: Berrett-Koehler Publishers Inc.
- knowledge base., 2015. *Threats to Construct Validity* [online]. Available from: <http://www.socialresearchmethods.net/kb/consthre.php> [accessed 25 October 2015].
- Krathwohl, D.R., 2002. A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, 41 (4), 212.
- Kwakman, K., 1999. *Leren van docenten tijdens de beroepsloopbaan. Studies naar professionalisering op de werkplek in het voortgezet onderwijs*. Academisch proefschrift. K.U. Nijmegen.
- Kwakman, C.E., 2001. Leren van professionals: Onderzoek naar leren-op-de-werkplek van docenten. Professional learning: Research into teachers' learning at the workplace. *Pedagogische Studiën*, 78 (2), 106-119.
- Kwakman, K., 2003. Factors affecting teachers' participation in professional learning activities. *Teaching And Teacher Education*, 19 (2), 149-170.
- Marzano, R.J., and Kendall, J.S., 2007. *The new taxonomy of educational objectives*. 2nd ed. California: Corwin Press.

- McDaniel, O., et al., 2010. *Op de grens van individuele verantwoordelijkheid en bestuurlijk verlangen*. Woerden: MBO raad.
- McDaniel, O., et al., 2009. *De professionaliteit van MBO-docenten in vergelijkend perspectief.gids voor BVE*. Den Haag: Reed Business.
- Misra, P., 2011. VET teachers in Europe: policies, practices and challenges. *Journal Of Vocational Education And Training*, 63 (1), 27-45.
- Nieuwenhuis, L.F.M., 2012. *Leven lang leren on the roc's, een visie op werken en leren in het MBO*. Heerlen: open universiteit.
- OECD (Organisation for Economic Cooperation and Development), 2005. *Teachers matter:attracting, developing and retaining effective teachers*. Paris: OECD.
- OECD, 2009. *Learning for jobs. OECD Policy Review of Vocational Education and Training (initial report)*. Paris: OECD.
- Opfer, D., and Pedder, D., 2011. Conceptualizing teacher professional learning. *Review of Educational Research*, 3 (81), 376-407.
- Oudeman, M., 2010. *Naar meer focus op het mbo. Advies van de Commissie Onderwijs en Besturing BVE*. Den Haag: BE.
- Plomp, T., and Nieveen, N., ed., 2010. *An introduction to educational design research*. 3rd print. Enschede: SLO.
- Nieveen, N., 2010. Formative Evaluation in Educational Design Research. In: Plomp, T., and Nieveen, N., eds. *An introduction to educational design research*. 3rd print. Enschede: SLO, 89-101.
- Sachs, J., and Mockler, N., 2012. Performance cultures of teaching: threat or opportunity?. In: C. Day, ed. *The routledge international handbook of teacher and school development*. New York: Routledge, 33-43.
- Saunders, R., 2012. Assessment of Professional Development for Teachers in the Vocational Education and Training Sector: An Examination of the Concerns Based Adoption Model. *Australian Journal Of Education*, 56 (2), 182-204.
- Schwarzer, R., and Jerusalem, M., 1995. Generalized Self-Efficacy scale. In: J. Weinman, S.Wright and M. Johnston, *Measures in health psychology: A user's portfolio. Causal and control beliefs*. Windsor, England: NFER-NELSON, 35-37.
- Schwarzer, R., and Hallum, S., 2008. Perceived Teacher Self-Efficacy as a Predictor of Job Stress and Burnout: Mediation Analyses. *Applied Psychology: An International Review*, 57, 152-171.
- Schwarzer, R., Schmitz, G.S. and Daytner, G.T., 1999. *teacher self-efficacy* [online]. Available from: http://userpage.fu-berlin.de/~health/teacher_se.htm [accessed 25 May 2012].
- Stamos, 2014. *market information educational sectors* [online]. Available from: <http://www.stamos.nl/index.rfx?verb=showitem&item=6.2.16&view=table> [accessed 25 October 2014].
- Smith, C., and Gillepsie, M., 2007. Research on professional development and teacher change: Implications for adult basic education. *Review of Adult Learning and Literacy*, 7, 205-244.
- Steiner, L., 2004. *Designing effective professional development experiences: what do we know?*. 2nd ed. Illinois: Naperville.
- Stevens, J.P., 1992. *thresholds used for factor loading cut-offs?* [online]. Cambridge, Cognition and Brain Sciences Unit. Available from: <http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/thresholds> [accessed October 4, 2015].

- Thurlings, M., Evers, A.T., and Vermeulen, M., 2015. Toward a Model of Explaining Teachers' Innovative Behaviour: A Literature Review. *Review Of Educational Research*, 85 (3), 430-471.
- Timperley, H., 2008. *Teacher professional learning and development*. Belly: Imprimerie Nouvelle Gonnet.
- Van den Berg, R., and Vandenberghe, R., 1981. *Onderwijsinnovatie in verschuivend perspectief*. Tilburg: Zwijsen.
- Veen, van K., et al., 2010. *Professional development of teachers*. Leiden: ICLON/VU.
- Verbeek, F., et al., 2009. *Professionalisering in de BVE-sector, Ontwikkeling van organisatie en personeel*. Amsterdam: SC-Kohnstamm instituut.
- Verbiest, E., and Timmerman, M., 2008. Naar duurzame schoolontwikkeling. In: Verbiest, E., ed. *Scholen duurzaam ontwikkelen. Bouwen aan professionele leergemeenschappen*. Antwerpen/Apeldoorn: Garant, 189-200.
- Villegas-Reimers, E., 2003. *Teacher professional development- an international review of the literature*. Paris: International Institute for Educational Planning.
- Volmari, K., Helakorpi, S., and Frimodt, R., eds., 2009. *Competence framework for VET professions, Handbook for practioners*. Sastamala: CEDEFOP handbook for practitioners
- Wayne, A.J., et al., 2008. Experimenting with Teacher Professional Development: Motives and Methods. *Educational Researcher*, 37 (8), 469-479.
-
- Zepeda, S.J., 2012. *Professional development, what works*. 2nd ed .Larchmont. NY: Eye on education.